Considerations in Cost of Electronic Document Storage

Stewart Bruner

ERR&D Committee

July 30, 2013

On a quest for numbers…

Understanding Storage Hardware

- A terabyte of storage purchased for home is <u>not</u> equivalent to a terabyte of storage purchased for business:
  - Home <u>backs up</u> primary disk inside the computer
  - AOC use <u>IS</u> primary disk
  - Generic ACE Hardware bolt versus Aerospace "Grade 8" alloy bolt
- AOC uses RAID 10 to meet 1-503(D)(2)(c). It enables replacement of failed drive without any data loss, but requires 4 drives instead of 1.
  - Home $99.99 solution has MTBF=1.6M hrs and AFR of .55%
  - A failure at home means only inconvenience and installation of another non-RAID drive
  - RAID 10 drives are set up as mirrored pairs and then striped
  - AOC ~$350 solution has MTBF=6.4M hrs with AFR of .1375%.
  - A failure of a lone drive at work means loss of public records and court/justice partner/business access to those records
  - Failure of a RAID 10 drive means work continues while a new drive is inserted, remirrored from the working drive in the pair

Data Storage Macro View 2003 to 2013

2003 Storage Condition

- *Storage Technology*: FiberChannel & AIX SAN
- *Cost of Storage:* ~$240K (h/w only)
- *Amount of Data:* 716 GB (total capacity)
- *Floorspace Needed:* 1 rack

2013 Storage Condition

- *Storage Technology*: HP 3Par StoreServ (+ 6 previous technologies SANs not converted)
- *Cost of Storage:* $1.84M (incl. 550GB LJ EDMS storage)
- *Amount of Data:* 76,895 GB
- *Floorspace Needed:* 6 racks total in 2 facilities
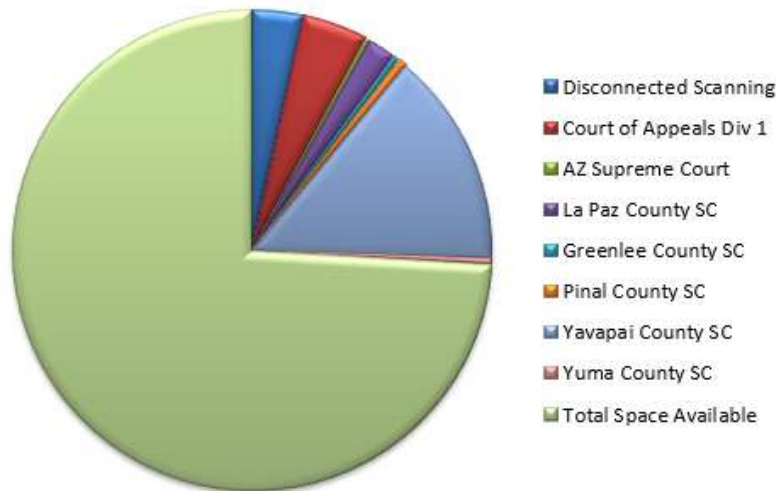- Cost per GB 10X less than in 2003 + faster access

2013 Case Document Storage Details
2003 = ZERO (all on local systems)

TOTAL SPACE USED (GB)	1461
Disconnected Scanning	192.10
Court of Appeals Div 1	246.88
AZ Supreme Court	20.16
La Paz County SC	106.79
Greenlee County SC	22.42
Pinal County SC	34.88
Yavapai County SC	814.80
Yuma County SC	23.42
Total Space Available	4,178

Total Docs in System	4,194,598

Average Doc Size (KB)	333

Pie chart legend:
- Disconnected Scanning
- Court of Appeals Div 1
- AZ Supreme Court
- La Paz County SC
- Greenlee County SC
- Pinal County SC
- Yavapai County SC
- Yuma County SC
- Total Space Available

Central Document Repository

- *Storage Details*: SQL DES SAN (incl. some other systems)
- *Cost of Storage:* $236K (h/w + ½ of SAN director)
- *Total Capacity:* 4,178 GB
- *Amount Used:* 1,461 GB
- *Amount Remaining:* 2717 GB
- *Avg Doc Size:* 333 KB
- *Total Docs:* 4.2M

Large size to accommodate GJ historical document load from counties

2013 Case Document Storage Details
2003 = ZERO (program didn't exist yet)

TOTALS			
DOCS	USED (GB)	AVAILABLE (GB)	AVG. DOC SIZE (KB)
2,805,441	192.10	557	58

USED (GB) 26%

AVAILABLE (GB) 74%
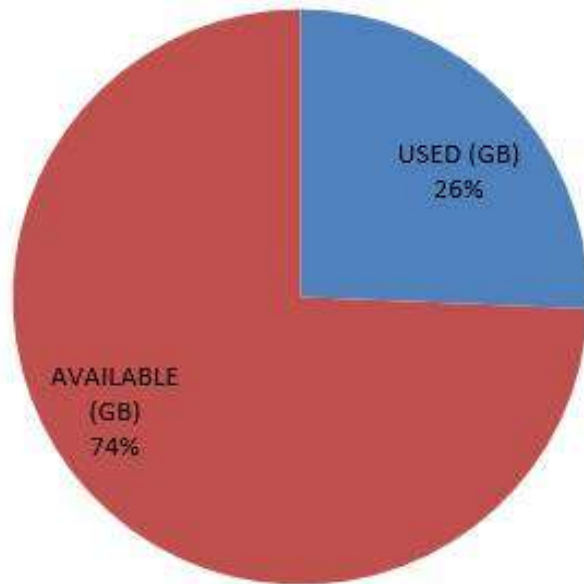
LJ Shared EDMS

- *Storage Details*: SQL SAN (incl <u>many</u> other systems)
- *Cost of Storage:* ~$243K (h/w only, incl. data and docs)
- *Total Doc Capacity:* 557 GB
- *Amount Used:* 192 GB
- *Amount Remaining:* 365 GB
- *Avg Doc Size*: 58 KB
- *Total Docs:* 2.8M

Small size predicated on LJ retention schedule enforcement

But what if…

POLICY CHANGE CAUSES AN INCREASE IN NUMBER OF DOCUMENTS STORED VS DESIGNED AMOUNT…?
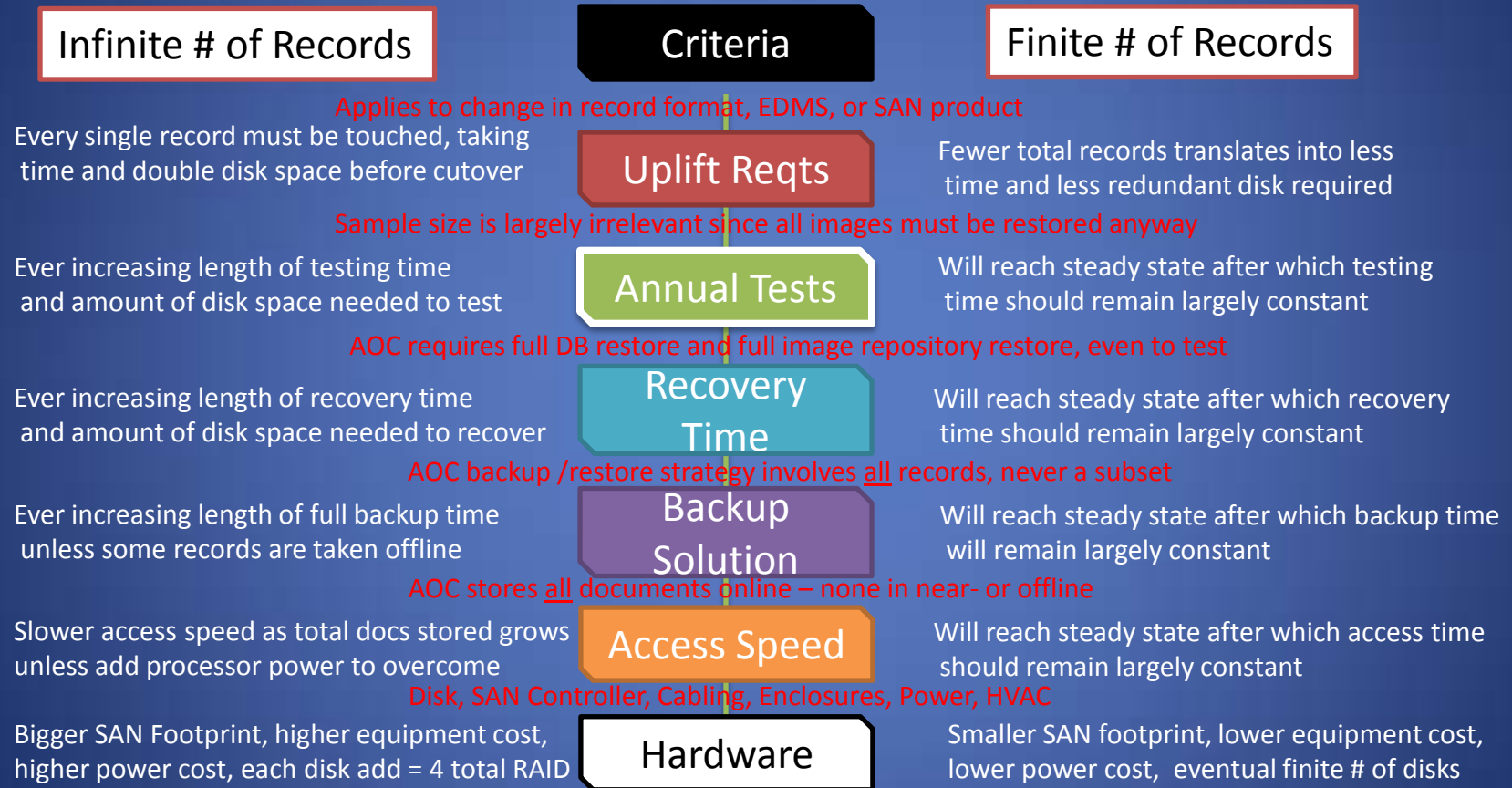
Common Foundational Items

Always necessary, but amount and/or size will vary by the total amount of data and documents being stored

- Includes floorspace, physical and logical security, servers, environmental controls, operating system, database mgt system, EDMS application, backup software, vendor support/maintenance contracts, disaster recovery strategy, support personnel, management, documentation, testing resources
- ACJA 1-506(G)(3) makes GJ clerk's local EDMS official record – same general requirements exist on those systems as on AOC systems
  - 12-284.01 funds to care for local systems, not state
- More total records online means more risk of PII loss in the event of a breach at any location (pre-2010 worse)

Comparison of Factors Related to Storage Costs

NOTE: A terabyte of storage purchased for home is not equivalent to a terabyte of storage purchased for business. Home backs up primary disk inside the computer. AOC uses RAID 10 for redundancy, requiring 4 drives instead of 1. Home $100 solution with an MTBF=1.6M hrs and AFR of .55% is business solution of $350 and MTBF=6.4M hrs with AFR of .1375%. A failure at home means only inconvenience while a failure at work means loss of public records and court/justice partner /business access to those records!

Infinite # of Records	Criteria	Finite # of Records
	Applies to change in record format, EDMS, or SAN product	
Every single record must be touched, taking time and double disk space before cutover	**Uplift Reqts**	Fewer total records translates into less time and less redundant disk required
	Sample size is largely irrelevant since all images must be restored anyway	
Ever increasing length of testing time and amount of disk space needed to test	**Annual Tests**	Will reach steady state after which testing time should remain largely constant
	AOC requires full DB restore and full image repository restore, even to test	
Ever increasing length of recovery time and amount of disk space needed to recover	**Recovery Time**	Will reach steady state after which recovery time should remain largely constant
	AOC backup /restore strategy involves all records, never a subset	
Ever increasing length of full backup time unless some records are taken offline	**Backup Solution**	Will reach steady state after which backup time will remain largely constant
	AOC stores all documents online – none in near- or offline	
Slower access speed as total docs stored grows unless add processor power to overcome	**Access Speed**	Will reach steady state after which access time should remain largely constant
	Disk, SAN Controller, Cabling, Enclosures, Power, HVAC	
Bigger SAN Footprint, higher equipment cost, higher power cost, each disk add = 4 total RAID	**Hardware**	Smaller SAN footprint, lower equipment cost, lower power cost, eventual finite # of disks

Foundational items include security, servers, operating systems, database mgt system, EDMS application, support personnel, management, documentation, vendor support/maintenance contracts, backup software, testing resources

1. Storage Hardware Comparison

Infinite # of Records

- Each disk added = 4 total disks RAID 10 primary storage
- Holding SAN footprint requires costly new technology to be infused regularly
- More floorspace = more HVAC = more power $$
- Ever-increasing equipment and maint cost
- Ever increasing energy costs

Finite # of Records

- On average, records are removed at same rate added
- SAN footprint eventually reaches a maximum and holds
- Equipment cost capped
- Energy cost capped

2. Access Speed Considerations

Infinite # of Records

- Access speed of any single record slows as total number of records online increases

- Overcoming the slowness requires increased processor power or

- Some records must be moved offline to speed access to other records

Finite # of Records

- Access speed of any single record will slow over time then reach steady state

- Will remain constant once maximum number of records is reached

- Only requires sufficient processor power to address records remaining online

3. Backup Solution Considerations

Infinite # of Records

- Time to complete a full backup will grow until insufficient number of hours exist to finish over night then over weekend
- Will require some records to be taken offline to shorten full backup time to off hours
- Offsite media needs continue to grow as does cost to maintain offsite

Finite # of Records

- Backup time will reach a maximum length after which will remain largely constant
- Less likely any records need to be taken offline to complete backup off hours
- Offsite storage costs eventually level off at steady state

4. Recovery Time Considerations

Infinite # of Records

- Ever increasing amount of disk space needed to write recovered files <u>back</u> onto
- Ever increasing length of time needed to perform recovery, lengthening minimum outage time
- Increased amount of computing resources need to be in place before performing recovery

Finite # of Records

- Will reach steady state after which recovery time and needs will remain largely constant and easy to quantify

5. ASLAPR Annual Testing Considerations (1/1/14)

Infinite # of Records

- Annual inspections must be performed on statistically significant sample – sample increases as amount increases

- Ever increasing length of testing time to meet F.1 and F.2 test requirements each year

- Ever increasing amount of disk space needed to perform F.1 and F.2 testing each year

Finite # of Records

- Record amount will reach steady state after which sample size, testing time, and space needed to meet F.1 and F.2 will remain largely constant

6. Upgrade/Product Change Considerations

Infinite # of Records

- Every single record must be touched in conversion; more records means longer conversion time and increased project complexity

- Old records remain while conversion is tested – instantly doubling disk space needed

Finite # of records

- Fewer total records translates into less time to convert and less redundant disk space nee

- Enables more product options and lower cost to convert.

7. Full-Text Searching Considerations

Infinite # of Records

- All scanned records ever created must be indexed in separate OCR process and remain online

- Search index grows infinitely; reaches limitation of product used

- Query time is function of total items in index, so access slows over time

Finite # of Records

- Lesser number of records subject to OCR and online

- Search index reaches steady-state size as items are removed at same rate as addition

- Less likely to outgrow product used

- Faster query time to get accurate result

8. Redaction Time & Cost Considerations

Infinite # of Records

- All records kept are subject to access therefore redaction
- All records must be online to be redacted, so access time slows to queue for redaction process
- For safety, record must be reviewed, even if automated redaction; more records to review equals more human effort

Finite # of Records

- Lesser number of records subject to redaction
- Faster access time to get to process
- More likely to characterize rules for automated redaction

The Takeway...

I.T. IS POISED TO CARRY OUT THE POLICIES DECIDED – THERE IS ALWAYS A COST, HOWEVER